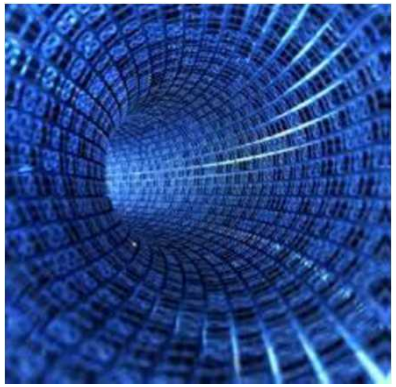
SAND2015-1930PE

Exceptional service in the national interest

Sandia National Laboratories



Finding the Digital Equivalent of a Needle in a Haystack

Corey Reitz

U.S. DEPARTMENT OF ENERGY

NNSA National Nuclear Security Administration

Sandia National Laboratories is a multi-program laboratory managed and operated by Sandia Corporation, a wholly owned subsidiary of Lockheed Martin Corporation, for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-AC04-94AL85000. SAND NO. 2011-XXXXP

A little about me…

Sandia National Laboratories

- Principal Software Systems Engineer at Sandia National Laboratories
- Formal Education
  - B.B.A., M.B.A., & J.D. all from the University of New Mexico
- Working with information technology since 1998.
- Licensed attorney in New Mexico since 2009.
- Focused almost exclusively on e-discovery, WFA, and legal technology since 2008.
- Enjoy dealing with the combination of challenges posed by technological and legal advances.

E-Discovery Basics

Sandia National Laboratories

- Discovery - The pretrial phase in litigation for obtaining evidence from an opposing party. Goal: Bring facts to light so the attorneys on both sides can determine if there are merits and/or defenses to claims and create a legal strategy (drop, settle, or proceed to trial).
  - Traditional Discovery: Paper based
  - E-Discovery: electronically stored information (e-mail, word processor files, etc.)
- Sources of Rules
  - Federal Rules of Civil Procedure- amended in 2006 to formally address electronically stored information (ESI)
  - Federal Rules of Evidence
  - State rules of Civil Procedure & Evidence
  - Federal and State Common law

E-Discovery Challenges

Sandia National Laboratories

- Ensuring evidence is not altered or deleted
- Constantly changing technology
- Too much data, impossible to manually review all case files in a timely manner
  - 1 GB= 70,000-80,000 text pages or 35-40 banker boxes
- Expensive to review all files
  - Document reviewers are expensive
- Bottom line: Need tools and new approaches to adequately leverage e-discovery in legal cases to identify relevant information.

Consequences of Mismanaging E-Discovery

Sandia National Laboratories

- Mismanagement of E-Discovery
  - Evasive, Incomplete, or complete failure to disclose or answer on time
  - Spoliation- "the destruction or significant alteration of evidence, or the failure to preserve property for another's use as evidence in pending or reasonably foreseeable litigation." *Cache La Poudre, LLC v. Land O'Lakes, Inc*., 244 F.R.D. 614, 613 (D.Colo.2007)
- Sanctions for not obeying a court order or Spoliation
  - Paying other parties expenses and attorney's fees
    - *Qualcomm Inc. v. Broadcom Corp. (Qualcomm I)*, No. 05cv1958-B (BLM), 2008 WL 66932 (S.D. Cal. Jan. 7, 2008).
  - Adverse instructions to the fact-finder
  - Dismissing the action (Plaintiff) or Default Judgment (Defendant)
- Negative publicity

Follow the Rules to Avoid Sanctions

Sandia National Laboratories

- 26(g) Signing Disclosures and Discovery Requests, Responses, and Objections.
  - (1) *Signature Required; Effect of Signature.* Every disclosure under Rule 26(a)(1) or (a)(3) and every discovery request, response, or objection must be signed by at least one attorney of record in the attorney's own name—or by the party personally, if unrepresented—and must state the signer's address, e-mail address, and telephone number. By signing, an attorney or party certifies that to the best of the person's knowledge, information, and belief formed after a reasonable inquiry:
    - (A) with respect to a disclosure, it is complete and correct as of the time it is made

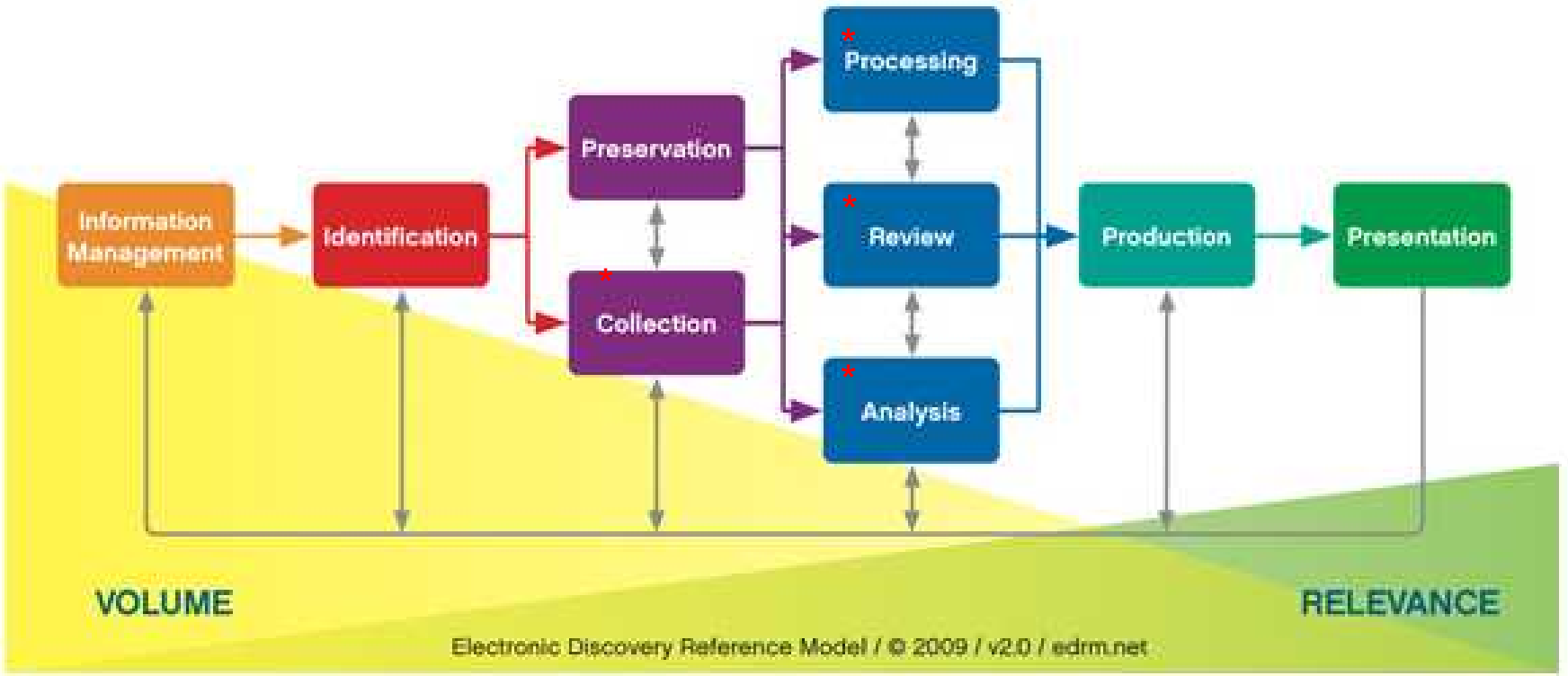
What is a reasonable search/inquiry?

Sandia National Laboratories

- Attorneys may not know for certain, but they have some touch points:
  - Parties don't have to produce information that is "not reasonably accessible because of undue burden or cost" FED R. CIV. P. 26(b)(2)(B).
  - "The Standard for Production of ESI is not Perfection" *Chen-Oster v. Goldman, Sachs & Co.*, 285 F.R.D. 294, 306 (S.D.N.Y. 2012).
  - The "reasonable inquiry" requirement is met where the investigation undertaken is "reasonable under the circumstances;" it is an "objective standard similar to the one imposed by Rule 11" Advisory Committee Notes, FED. R. CIV. P. 26(g);
- Following best practices will help to strengthen the argument that an individual is performing a reasonable search/inquiry.
  - The Sedona Principles (see https://thesedonaconference.org)
  - The E-Discovery Reference Model (see http://www.edrm.net)

Sandia National Laboratories

Electronic Discovery Reference Model



Identification

- There is no substitute for communicating with the people involved in the matter to inform them of the type(s) of information that will be relevant and to ask them where that information may be stored.
  - Desktop/laptop computers, network shares, e-mail servers, enterprise applications (HR, Accounting, Benefit systems, document management), databases, mobile devices, SharePoint Sites, external media (USB drives), social media sites, file cabinets, etc.
- Some cases are large and it is difficult to discuss the matter with all of those who are involved. In these cases, software can help.
  - Legal hold software to issue notices, send reminders, and send on-line interviews to solicit information from the individuals and track them.
    - Examples: Relativity Legal Hold, Exterro Legal Hold, Encase Legal Hold, Bridgeway Legal Hold, HP Legal Hold (HP Autonomy), etc.

Collection

Sandia National Laboratories

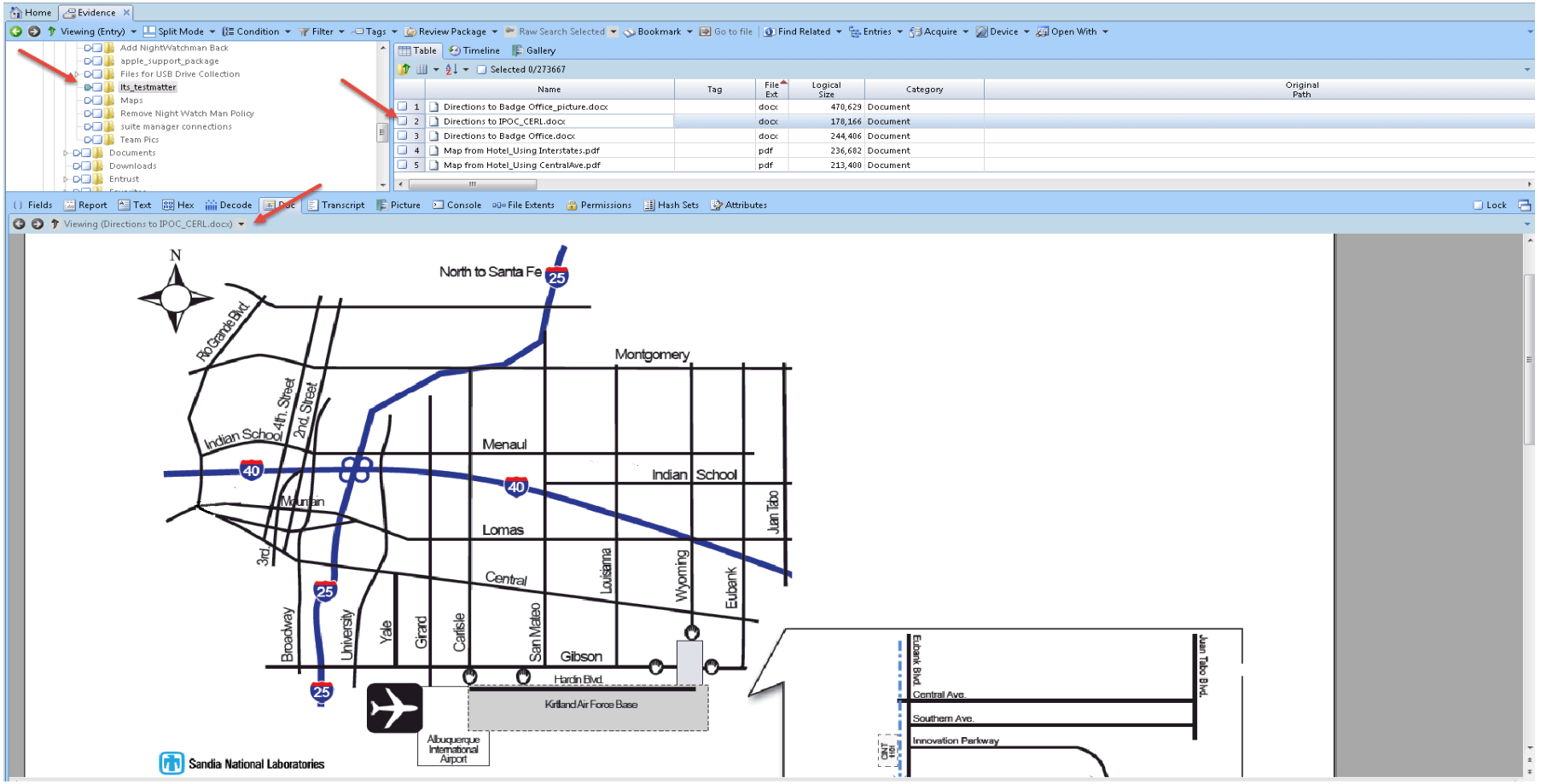
- Collecting on behalf of individuals v. self-collection
  - Collections on behalf of individuals will be consistent and quantitative, Self-collection will likely be more tailored, but less consistent.
- Focusing on collections on behalf of individuals
  - Physical versus Logical collections
  - Ideally the collector of the data doesn't want to go back to the well…Doesn't want too much data, doesn't want too little data…wants it just right just like Goldilocks.
- Remotely access the device or have the device brought to you?
  - If remote access, issues with the network (down times or latency), computer being taken off the network (e.g. laptops when user goes home), issues with agent not running correctly on the computer, etc.
  - Downtime to the user if the device is brought in, but no network issues
- Various OS and file systems. May need to collect using different criteria to accommodate for the differences between OS and file systems (e.g. Mac OS X vs. Windows 7).
- Various encryption mechanisms….Full disk encryption (FDE), various software encryption (e.g. Credant, FileVault, Bitlocker), etc.

Collection Tools

Sandia National Laboratories

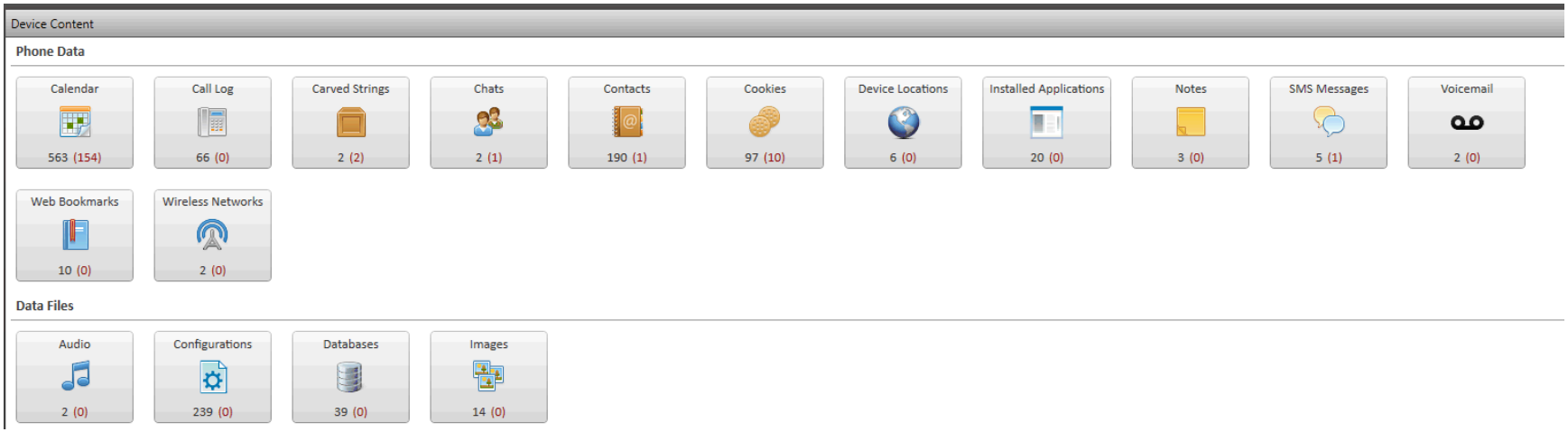
- Enterprise collection tools allow you to collect data from machines that are running an agent remotely.
  - Examples: FTK, Encase, Exterro, HP eDiscovery, Xerox ViewPoint
  - Many can handle encryption keys to collect from encrypted machines.
  - Challenges: Machine disconnects from the network due to mandatory patches or turned off by owner at the end of the day, Machine shuts down due to energy saving software, etc.
- Portable devices (e.g. USB keys) that have criteria on them to collect from those computers that are on slower connections (e.g. VPN)
- Mobile device collection tools
  - Examples: Paraben Device Seizure, UFED Cellebrite, iPhone Analyzer
- **Test** to make sure your tools work the way you think they do.
- With all this attention on electronically stored information, don’t forget to collect paper-based evidence!

Collection Tools



Collection Tools

- Mobile Forensics- capture data such as calendar entries, call log, IM, notes, voicemail and other audio files, web history, GPS data, and wireless networks that have been accessed.



Processing

Sandia National Laboratories

- E-mail- various types (Exchange/Outlook, Novell, MBOX, etc.)
- Non text based files (pictures, audio, video), non-searchable PDFs...isolate and OCR?
- Encryption, password protected files
- OLE objects- treat them as separate files or as part of parent?
- Compound files- pst, zip, embedded files
  - Expand and extract
- Nuances of text searches (being over/under inclusive)...advice to perform statistical sampling with an agreed upon CI
- Create good criteria- keywords, file filtering, etc.
  - Wolfram Alpha- "words containing <word>"
- Know the strength and limitations of your tools by testing (each product, each version)
  - Ex: html based e-mail not indexing all of the text

De-duplication

Sandia National Laboratories

- Method for reducing the volume of files
- Files- using a cryptographic hash function (commonly MD5 or SHA-1) the processing application creates the equivalent of a unique digital finger print for each file by looking at the bytes. If the finger prints match, the files are most likely duplicates.
  - Example of an MD5 hash value: 58057225245fv1d26g4bc422per57eb4
- E-mail- Takes the text from various metadata fields such as BCC, Body, CC, From, Subject, To, Attach, and puts them together in one long string of text. Then the software creates a hash of that value and compares it to other e-mail messages.
- De-duplication can be performed either globally or on a per custodian basis.

Create an Index & Search

Sandia National Laboratories

- Index- stores data in a format that facilitates quick information retrieval
- Ensure you know what words are “noise” words for the indexing engine, these will be ignored when creating the index. Examples: to, and, the, a, is, at, on, which
  - Original text: The quick brown fox jumped over the lazy dog
  - Index: Quick brown fox jumped over lazy dog
- Searches (each tool may have its own syntax)
  - Phrase: “corey likes running”
  - Key word: running
  - Proximity: corey w/5 running
  - Wildcard: r?n
  - Stemming: run~

Document	Words
Document A	Corey, likes Running,
Document B	Corey, runs
Document C	Water, running

Processing Tools

Sandia National Laboratories

- Processing tools allow you to filter down the data by using various criteria (keywords, date ranges, file types, file location, etc.)
- Examples: FTK, Encase, LexisNexis Concordance, Recommind Axcelerate

Processing Tools

Sandia National Laboratories

The following folder list will be included:
\users\csreitz
Documents will be included if they were Created, Accessed or Modified
After: 1/10/2015 12:00:00 AM and Before: 3/10/2015 11:59:59 PM
Documents will be excluded if they were Created, Accessed or Modified
After: 3/1/2015 12:00:00 AM and Before: 3/1/2015 11:59:59 PM
Documents with the following common file extension will be force included:
123, cal, gph, lss, slk, wk1, wk3, wk4, wkq, wks, wrk, xlb, xlk, xlm, xls, xlsx, xlt, xlw
The following index query terms with Minimum Word Length of 1 will be included:
corey likes running, corey w/2 reitz
The following document options will be performed:
Embedded documents will be included
Deleted files will be included
Compound files will be included

Review Considerations

Sandia National Laboratories

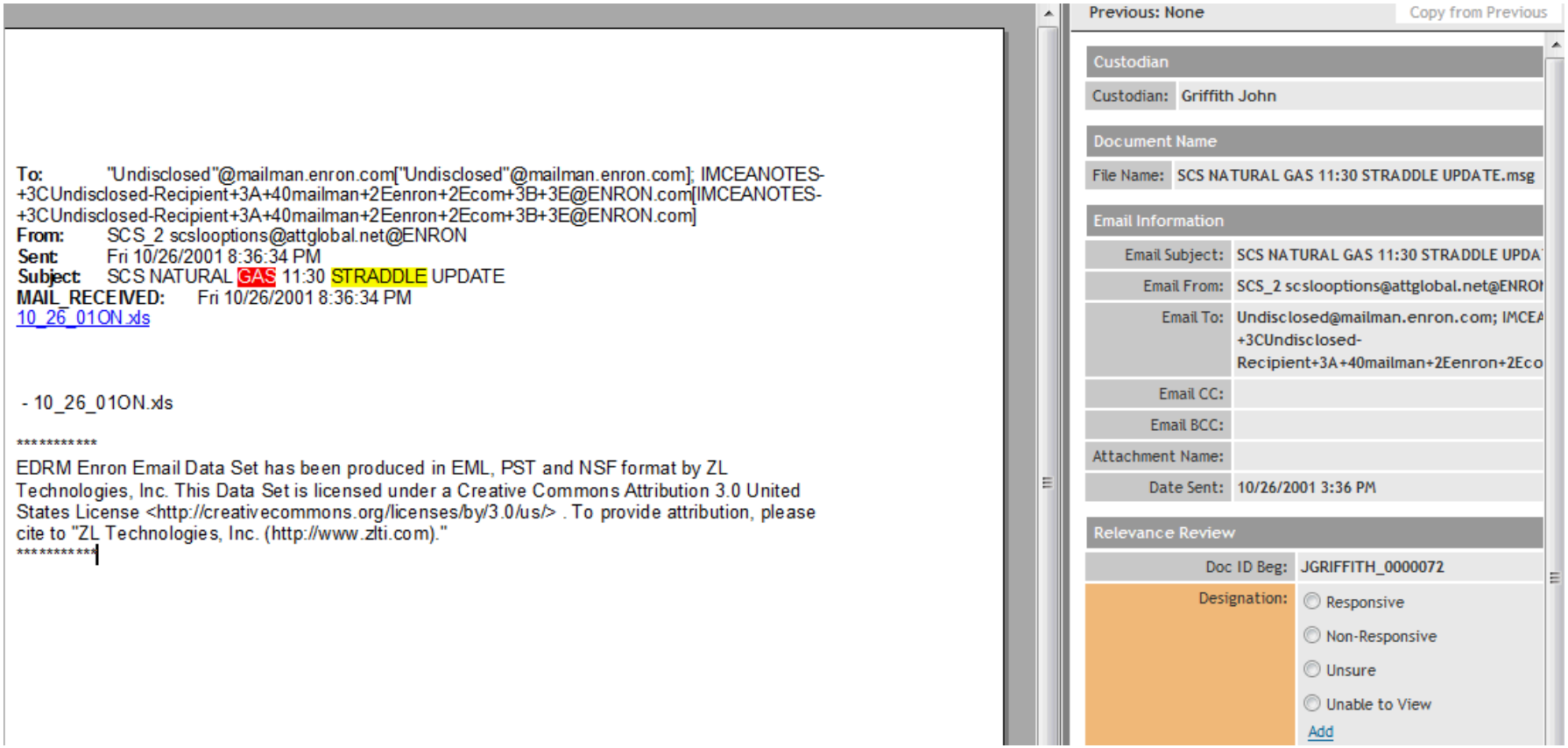
- Viewers allow you to view files that you don't have software for, but they generally don't render large files.
- Non-standard file formats...require the native software to review
- Some files are just better off being reviewed natively (e.g. Excel worksheets with formulas, audio files, video files)
- How much do you trust OCR? Should you review those files even if they are not responsive to key word searches?
- Use Review platform to keep reviewers from making mistakes
  - Example: Cannot tag a document as 'not-responsive' and 'responsive'
- Need to train document reviewers on what you are looking for and then sample their results and continuously train to your specifications.

Review Tools

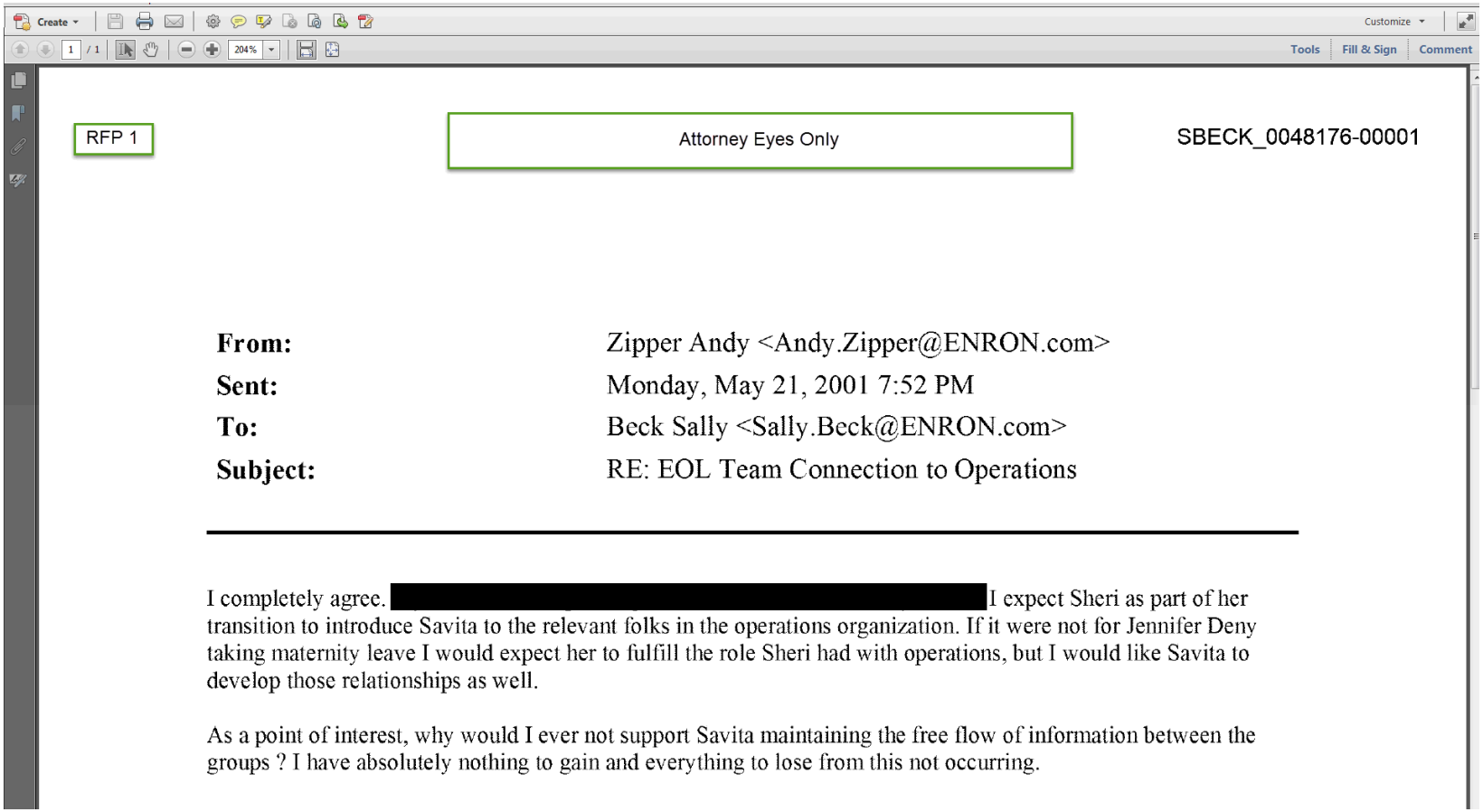
Sandia National Laboratories

- Review tools provide additional searching capabilities, allow for categorization of documents, e-mail threading, and provide viewers to be able to see files without having the native software installed on the reviewers machine.
- Review tools also allow the user to redact text, convert native files into tiff or pdf, and produce the results with labels.
- In very large matters, Review tools can be used to leverage computer assisted review techniques such as predictive coding.
- Examples: Concordance, Summation, Relativity, Symantec Ediscovery

Review Tools



Review Tools



RFP 1

Attorney Eyes Only

SBECK_0048176-00001

From: Zipper Andy <Andy.Zipper@ENRON.com>
Sent: Monday, May 21, 2001 7:52 PM
To: Beck Sally <Sally.Beck@ENRON.com>
Subject: RE: EOL Team Connection to Operations

I completely agree. [redacted] I expect Sheri as part of her transition to introduce Savita to the relevant folks in the operations organization. If it were not for Jennifer Deny taking maternity leave I would expect her to fulfill the role Sheri had with operations, but I would like Savita to develop those relationships as well.

As a point of interest, why would I ever not support Savita maintaining the free flow of information between the groups ? I have absolutely nothing to gain and everything to lose from this not occurring.

The Needle in the Haystack

Sandia National Laboratories

- Probability of finding the needle in a haystack: depends on how much hay you have to search through. ☺
- Probability of a file ending up on the evidence list for a case: "In e-discovery, only about .001 percent of [documents] even ends up making the evidence list." Ralph Losey, E-discovery Attorney
  - This means that only one of every 100,000 documents are likely to become admissible evidence in a legal dispute.
- Just for fun…by comparison the odds of being struck by lightning in your lifetime is one in 12,000 if you live to be 80 years old.

Valuable Skills in E-discovery

Sandia National Laboratories

- Solid understanding of information technology
  - Familiarity with common operating systems and file formats
  - Experience supporting software applications (patching, upgrading, etc.)
- Experience using E-discovery software tools
  - Collection, processing, and production tools
  - Document review platforms
- Familiar with the rules that govern e-discovery
  - Federal and state rules of civil procedure
  - Federal and state rules of evidence
  - Case law that interpret the rules
- Exposure to sound project management techniques
  - Project planning, risk assessments, continuous improvement, etc.

Job Outlook

Sandia National Laboratories

- Cowen Group 2012 Annual Salary Report Litigation Support
  - 71% of AmLaw 200 expected to hire additional staff in 2013
  - Litigation Support Analyst $69k-80k, Litigation Support Specialist $75k-95k, Litigation Support Project Manager $90k-125k, Regional eDiscovery Coordinator $120k-145k.
- ILTA's (International Legal Technology Association) 2012 IT Compensation Survey
  - ILTA surveys their membership annually for salary data
  - Average salaries: Litigation Support Analyst $72k, Litigation Support Specialist $71K, Litigation Support Supervisor $91k, Litigation Support Manager $121k.

Questions and Answers

Sandia National Laboratories